International Journal of Computer Science Engineering
and Information Technology Research (IJCSEITR)
ISSN(P): 2249-6831; ISSN(E): 2249-7943
Vol. 5, Issue 1, Feb 2015, 31-38

XVSGC TOOL OF QURABLE DATA COMPRESSION IN XML DATASETS

VIJAY GULHANE[1] & M. S. ALI[2]

[1]Sipna COET, Amravati, Maharashtra, India

[2]Rpof. Ram Meghe COEM, Badnera, Amravati, Maharashtra, India

ABSTRACT

This research describe in this paper is a new approach of partial query processing based on a SAX query parser and Query engine. It first loads the compressed XML document and then processes. The brief description of proposed scheme is presented, followed by the design. This begins with the introduction discussion experimental results, followed by its evaluation and compressions with some other existing technologies.

KEYWORDS: XML Parsing, XVSGC Approach, Data Compression, Partial Decompression, Quriable Compression

INTRODUCTION

Query processing plays a vital role in the database system. However little work has been done to study the effective compression and query processing in XML documents. Query processor refers to a collective set of strategies that processes the queries in some sort of time. The basic component of any query processor is the parser and the Query engine. Query engine attempt to resolve the query. Search engine and the Parser is heart of the query processor. Search engine searches the exact match for the incoming queries and then it optimize it for the result.

The compression time for XML database is calculated by running the same database For multiple times and average is taken for consideration of result. The main reason for doing this is to reduce the disk I/O influences on the results by loading the whole document into the physical memory.. Calculations of CR1 and CR2 are done using the following formulas-

The compression ratio is defined as follows

CR1=(Size of Compressed File *8)/Size Of Original File

CR2=1-(size of Compressed file /Size of Original File)*100

Compression Ratio Factor (CRF):- Normalize the Compression Ratio of XVSGC with Respect to XMill and XGRIND

CRF 1=CRXVSGC/XMILL

CRF 2=CRXVSGC/XGRIND

Compression Time Factor (CRT):- Normalize the Compression Time of XVSGC with Respect to XMill and XGRIND.

CRT1=CRTXVSGC/XMILL

CRT 2=CRTXVSGC/XGRIND

EXPERIMENTAL EVALUATION

The scheme of search engine, PARSER, proposed LZ base compressed query processing and studied for different load. The performance metrics use to mesure the performance of the query processor are minimum time taken to execute the result.

Query Response Time (QRT): Time Required T to execute the query

The results of the evaluation for the different document are shown in following table. And comparative graphs are shown.

Each query is fired more than 5 time for the same purpose. In above table we shows the five query and three different time period for the partial processing. At last we shows the average of that time as our result.

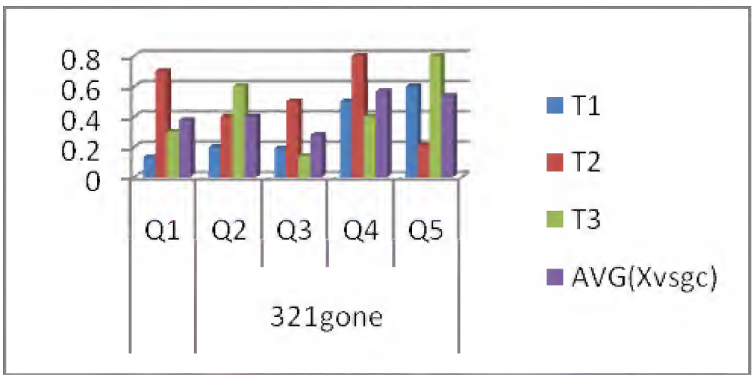

Figure 1: Qurey Response Time of XVSGC for 321gone

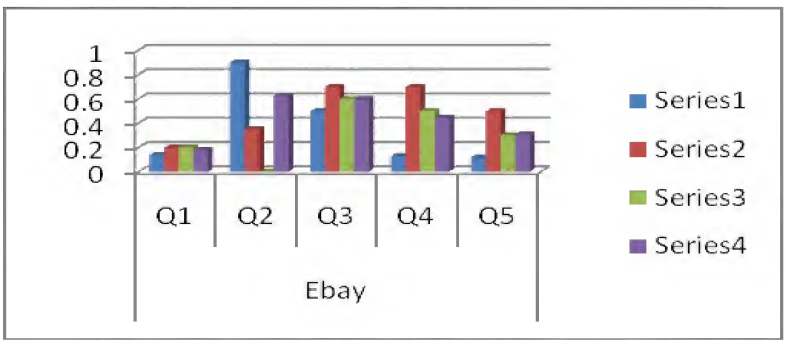

Figure 2: Qurey Responce Time of XVSGC for Ebay

Above figure 2 shows the graphical representation of the query response time for the XVSGC for Ebay. Graph is drawn using the our exprimental results.

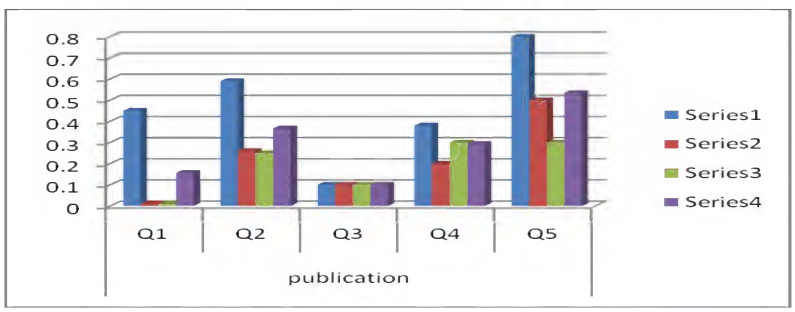

Figure 3: Qurey Response Time of XVSGC for DBLP (Publications)

Above figure 3 shows the graphical representation of the query responce time for the XVSGC for publication.

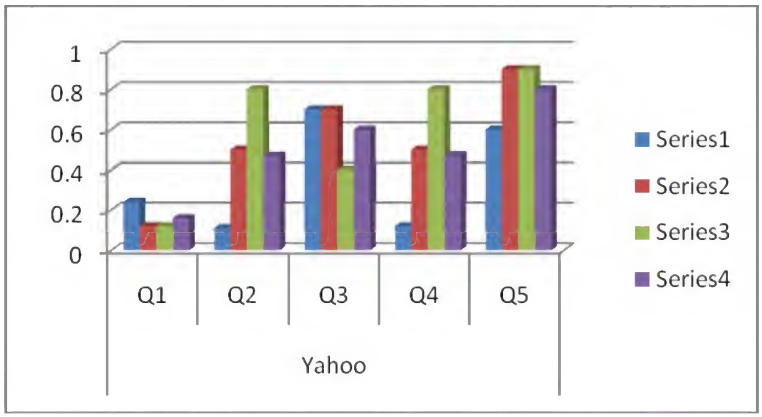

Figure 4: Qurey Response Time of XVSGC for Yahoo

Above figure 4 shows the graphical representation of the query res[ponce time for the XVSGC for Yahoo. Graph is drawn using the our exprimental results.

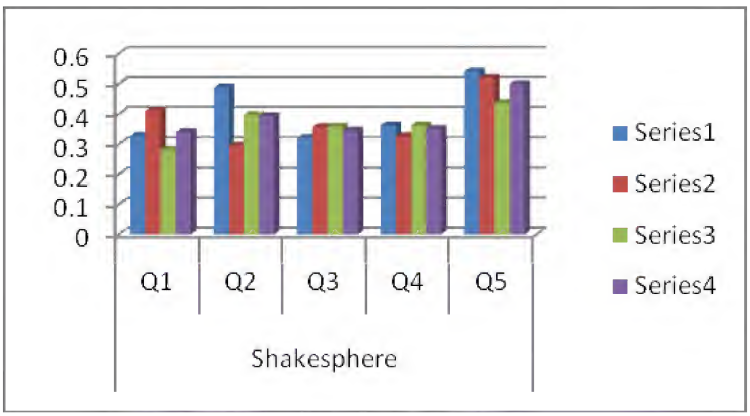

Figure 5: Qurey Response Time of XVSGC for Shakesphere

Above figure 5 shows the graphical representation of the query responce time for the XVSGC for Shakesphere. Graph is drawn using the our exprimental results.

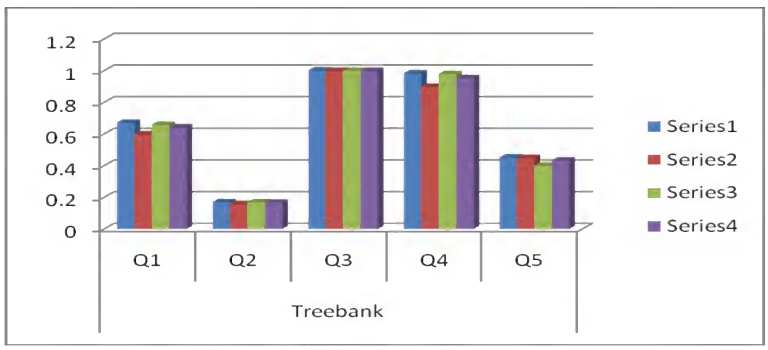

Figure 6: Qurey Response Time of XVSGC for Treebank

Above figure 6 shows the graphical representation of the query responce time for the XVSGC for Treebank. Graph is drawn using the our exprimental results.

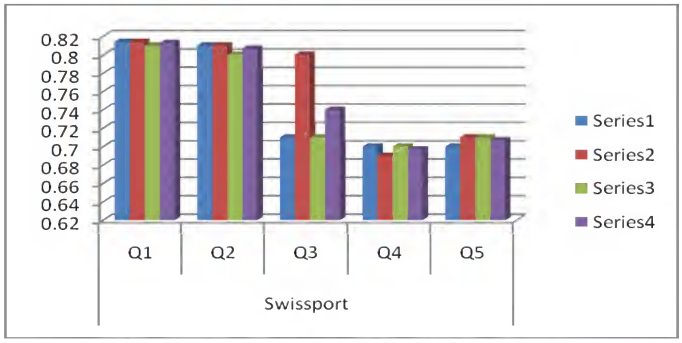

Figure 7: Qurey Response Time of XVSGC for Swissport

Above figure 7 shows the graphical representation of the query response time for the XVSGC for Swissport. Graph is drawn using the input from our experiments.

Table 1: Performance over Regular Documents

Name	Q. No	DT1 (in Sec)	DT2 (in Sec)	DT3 (in Sec)	AVG(Xvsgc) (in Sec)
Weblog	Q1	0.275	0.212	0.246	0.244333333
	Q2	0.114	0.245	0.168	0.175666667
	Q3	0.12	0.133	0.146	0.133
	Q4	0.185	0.226	0.252	0.221
	Q5	0.275	0.288	0.27	0.277666667

Table 2: Partial Query Performance of XVSGC

Name	O. Size	C. Size	Q. No	DT1 (in Sec)	DT2 (in Sec)	DT3 (in Sec)	AVG(Xvsgc) (in Sec)
321gone	24.9	22.5	Q1	0.131	0.7	0.3	0.377
			Q2	0.2	0.4	0.6	0.4
			Q3	0.19	0.5	0.14	0.276666667
			Q4	0.5	0.8	0.4	0.566666667
			Q5	0.6	0.21	0.8	0.536666667
Ebay	34.7	41.6	Q1	0.14	0.2	0.2	0.18
			Q2	0.9	0.35	.0.7	0.625
			Q3	0.5	0.7	0.6	0.6
			Q4	0.13	0.7	0.5	0.443333333
			Q5	0.12	0.5	0.3	0.306666667

Above Table 2 shows Query performance XVSGC over irregular document. (time for the XVSGC for 321gone and Ebay.)

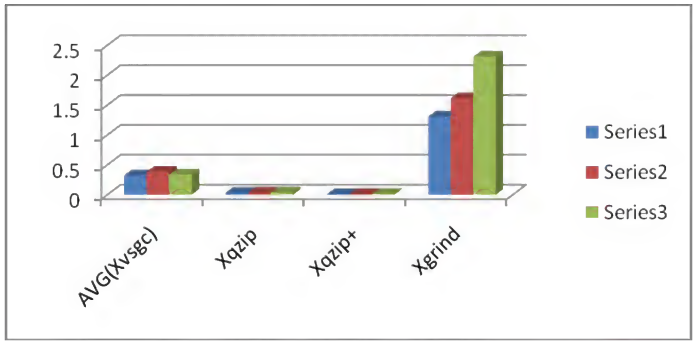

Figure 8: Query Performance OF XVSGC with Others of Textual Documents

Above figure 8 shows the Query response time for the different compressor. From the above figure we can conclude that Query response time of XVSGC is better than the XGRIND and the more than Xqzip and Xqzip+ as it is a non quarable compressor.

CONCLUSIONS

For this work, some of the XML compressor were examined and at the same time some query evaluation techniques are investigated with the objective of making efficient processing of XML Queries.

The proposed work intent to achieve 'optimal', efficient,'fast' query processing, and efficient query evaluation in XML compression.

Query processor plays an important role in the database management system. After evaluation it is found that our approach gives improved performance, Query response time of XVSGC is better than XGRIND and comparable to Xqzip and Xqzip+.

In wireless applications like PDA's, Smartphone's, Palmtops etc, non-querible compressor are not advantageous since querying data is important. From the results it is better choice to use XVSGC tool as it is giving better results than XGRIND.

Finally it can be concluded that using the proposed technique of compression of XML databases and query evaluation over it, will better improve the performance of XML database systems.

REFERENCES

1. Abramson, N. 1963. Information Theory and Coding. McGraw-Hill, New York.
2. AlHamadani, Baydaa "Retrieving Information from Compressed XML Documents According to Vague Queries" July, 2011 University of Huddersfield Repository http://eprints.hud.ac.uk/
3. Andrei Arion, Angela Bonifati, Ioana Manolescu, Andrea Pugliese "XQueC: A Query-Conscious Compressed XML Database" ACM Journal Name, Vol., No., 20, Pages 1–31.
4. Andrei Arion1, Angela Bonifati2, Gianni Costa2, Sandra D'Aguanno1, etel "Efficient Query Evaluation over Compressed XML Data" E. Bertino et al. (Eds.): EDBT 2004, LNCS 2992, pp. 200–218, 2004. _c Springer-Verlag Berlin Heidelberg 2004
5. Apostolico, A. and Fraenkel, A. S. 1985. Robust Transmission of Unbounded Strings Using Fibonacci Representations. Tech. Rep. CS85-14, Dept. of Appl. Math., The Weizmann Institute of Science, Rehovot, Sept.
6. Augeri, C. J., Bulutoglu, D. A., Mullins, B. E., Baldwin, R. O. & Leemon C. Baird, I. (2007). An analysis of XML compression efficiency. Proceedings of the 2007 workshop on Experimental computer science, ACM, San Diego, California.
7. Debra A. Lelewer and Daniel S. Hirschberg "Data Compression"
8. David Salomon, Data Compression: The Complete Reference, pub-SV, 2004.
9. Elias, P. 1987. Interval and Recency Rank Source Coding: Two On-Line Adaptive Variable-Length Schemes. IEEE Trans. Inform. Theory 33, 1 (Jan.), 3-10.

10. Faller, N. 1973. An Adaptive System for Data Compression. Record of the 7th Asilomar Conf. on Circuits, Systems and Computers (Pacific Grove, Ca., Nov.), 593-597.
11. G. Antoshenkov. Dictionary-Based Order-Preserving String Compression. VLDB Journal 6, page 26-39, (1997).
12. Gallager, R. G. 1978. Variations on a Theme by Huffman. IEEE Trans. Inform. Theory 24, 6 (Nov.), 668-674.
13. Gregory Leighton and Denilson Barbosa "Optimizing XML Compression (Extended Version)" arXiv:0905.4761v1 [cs. DB] 28 May 2009
14. G. Cleary, I. H. Witten, Data compression using adaptive coding and partial string matching, IEEE Trans. Commun. OM-32 (4) (1984) 396–402.
15. GZip Compressor, http://www.gzip.org/.
16. H. Liefke and D. Suciu. XMill: An Efficient Compressor for XML Data. Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 153-164 (2000).
17. Horspool, R. N. and Cormack, G. V. 1987. A Locally Adaptive Data Compression Scheme. Commun. ACM 16, 2 (Sept.), 792-794.
18. http://www.cs.washington.edu/research/xmldatasets/www/repository.html#SwissProt
19. Huffman, D. A. 1952. A Method for the Construction of Minimum-Redundancy Codes.
20. Proc. IRE 40, 9 (Sept.), 1098-1101.
21. J. Cheng and W. Ng. XQzip: Querying Compressed XML Using Structural Indexing.
22. Proceedings of EDBT (2004).
23. J. K. Min, M. J. Park, and C. W. Chung. XPRESS: A Queriable Compression for XML Data. Proceedings of the ACM SIGMOD International Conference on Management of Data (2003).
24. J. Clark, XML Path Language (XPath), 1999. http://www.w3.org/TR/xpath
25. J. Gailly and M. Adler, gzip1.2.4.http://www.gzip.org
26. James Cheng and Wilfred Ng "XQzip: Querying Compressed XML Using Structural Indexing" E. Bertino et al. (Eds.): EDBT 2004, LNCS 2992, pp. 219–236, 2004. _c Springer-Verlag Berlin Heidelberg 2004
27. JunKi Min MyungJae Park ChinWan Chung "XPRESS: A Queriable Compression for XML Data" SIGMOD 2003, June 912, 2003, San Diego, CA. Copyright 2003 ACM 158113634X/ 03/06 Knuth, D. E. 1985. Dynamic Huffman Coding. J. Algorithms 6, 2 (June), 163-180.
28. Llewellyn, J. A. 1987. Data Compression for a Source with Markov Characteristics. Computer J. 30, 2, 149-156.
29. M. Girardot, N. Sundaresan, Millau: An encoding format for efficient representation and exchange of XML over the Web, Comput. Networks 33 (1–6) (2000) 747–765. XAUST Compressor.
30. Michael Ley "DBLP — Some Lessons Learned" VLDB '09, August 2428, 2009, Lyon, France Copyright 2009 VLDB Endowment, ACM 0000000000000/ 00/00

31. Michael Ley "DBLP XML Requests- Appendix to the paper "DBLP — Some Lessons Learned" (June 17, 2009) Michael Ley Universit¨ at Trier, Informatik D–54286 Trier

APPENDICES

Query Response time for different XML document

Table 3

Name	Q. No	DT1 (in Sec)	DT2 (in Sec)	DT3 (in Sec)	AVG(Xvsgc) (in Sec)
321gone	Q1	0.131	0.7	0.3	0.377
	Q2	0.2	0.4	0.6	0.4
	Q3	0.19	0.5	0.14	0.276666667
	Q4	0.5	0.8	0.4	0.566666667
	Q5	0.6	0.21	0.8	0.536666667
Ebay	Q1	0.14	0.2	0.2	0.18
	Q2	0.9	0.35	.0.7	0.625
	Q3	0.5	0.7	0.6	0.6
	Q4	0.13	0.7	0.5	0.443333333
	Q5	0.12	0.5	0.3	0.306666667
Publication	Q1	0.45	0.01	0.01	0.156666667
	Q2	0.59	0.26	0.25	0.366666667
	Q3	0.1	0.1	0.1	0.1
	Q4	0.38	0.2	0.3	0.293333333
	Q5	0.8	0.5	0.3	0.533333333
Yahoo	Q1	0.24	0.12	0.12	0.16
	Q2	0.11	0.5	0.8	0.47
	Q3	0.7	0.7	0.4	0.6
	Q4	0.12	0.5	0.8	0.473333333
	Q5	0.6	0.9	0.9	0.8
Shakesphere	Q1	0.327	0.41	0.281	0.339333333
	Q2	0.487	0.295	0.396	0.392666667
	Q3	0.32	0.355	0.357	0.344
	Q4	0.361	0.326	0.361	0.349333333
	Q5	0.54	0.52	0.435	0.498333333
Treebank	Q1	0.674	0.6	0.66	0.644666667
	Q2	0.171	0.16	0.17	0.167
	Q3	1.003	1	1	1.001
	Q4	0.985	0.9	0.98	0.955
	Q5	0.453	0.45	0.4	0.434333333
Swissport	Q1	0.814	0.814	0.81	0.812666667
	Q2	0.81	0.81	0.8	0.806666667
	Q3	0.71	0.8	0.71	0.74
	Q4	0.7	0.69	0.7	0.696666667
	Q5	0.7	0.71	0.71	0.706666667
Weblog	Q1	0.275	0.212	0.246	0.244333333
	Q2	0.114	0.245	0.168	0.175666667
	Q3	0.12	0.133	0.146	0.133
	Q4	0.185	0.226	0.252	0.221
	Q5	0.275	0.288	0.27	0.277666667
Mondial	Q1	1.611	1.317	1.349	1.425666667
	Q2	0.6	0.7	0.5	0.6
	Q3	0.28	0.2	0.12	0.2
	Q4	0.24	0.12	0.12	0.16
	Q5	0.06	0.09	0.11	0.086666667

Above table shows the Query response time. In experiment we have taken more than five queries for the better result.